ABSTRACT OF THE DISCLOSURE

1   In order to align DNA sequence data traces, an experimental data trace

2   representing the positions of a first species of base within a target polynucleotide and a reference

3   data trace representing the positions of a second species of base (which may be the same as or

4   different from the first species) within a reference polynucleotide are obtained by separating

5   appropriate sequencing fragments generated from the target and reference polynucleotides on an

6   electrophoresis gel. For each reference data trace, a plurality of peaks corresponding to

7   fragments having a size in the range of 40 to 1200 bases are selected. A base number is assigned

8   to each of the selected peaks in the reference data trace, and a numerical "peak file" is created

9   with information about the peak number and migration time (or distance). This peak file is

10 analyzed to determine a set of polynomial coefficients which will allow substantial linearization

11 of a plot of peak number versus separation between adjacent peaks and alignment of the traces

12 with respect to each other. These coefficients are used to create a corrected time scale identifying

13 where peaks should be located on a given experimental gel. This corrected time scale is used to

14 guide the sampling of the experimental data, and for assignment of peaks within the data.